Course 6 Proj

Viraj Bhalala

April 23, 2017

Project Overview

In this project you will investigate the exponential distribution in R and compare it with the Central Limit Theorem. The exponential distribution can be simulated in R with rexp(n, lambda) where lambda is the rate parameter. The mean of exponential distribution is 1/lambda and the standard deviation is also 1/lambda. Set lambda = 0.2 for all of the simulations. You will investigate the distribution of averages of 40 exponentials. Note that you will need to do a thousand simulations.

Illustrate via simulation and associated explanatory text the properties of the distribution of the mean of 40 exponentials. You should

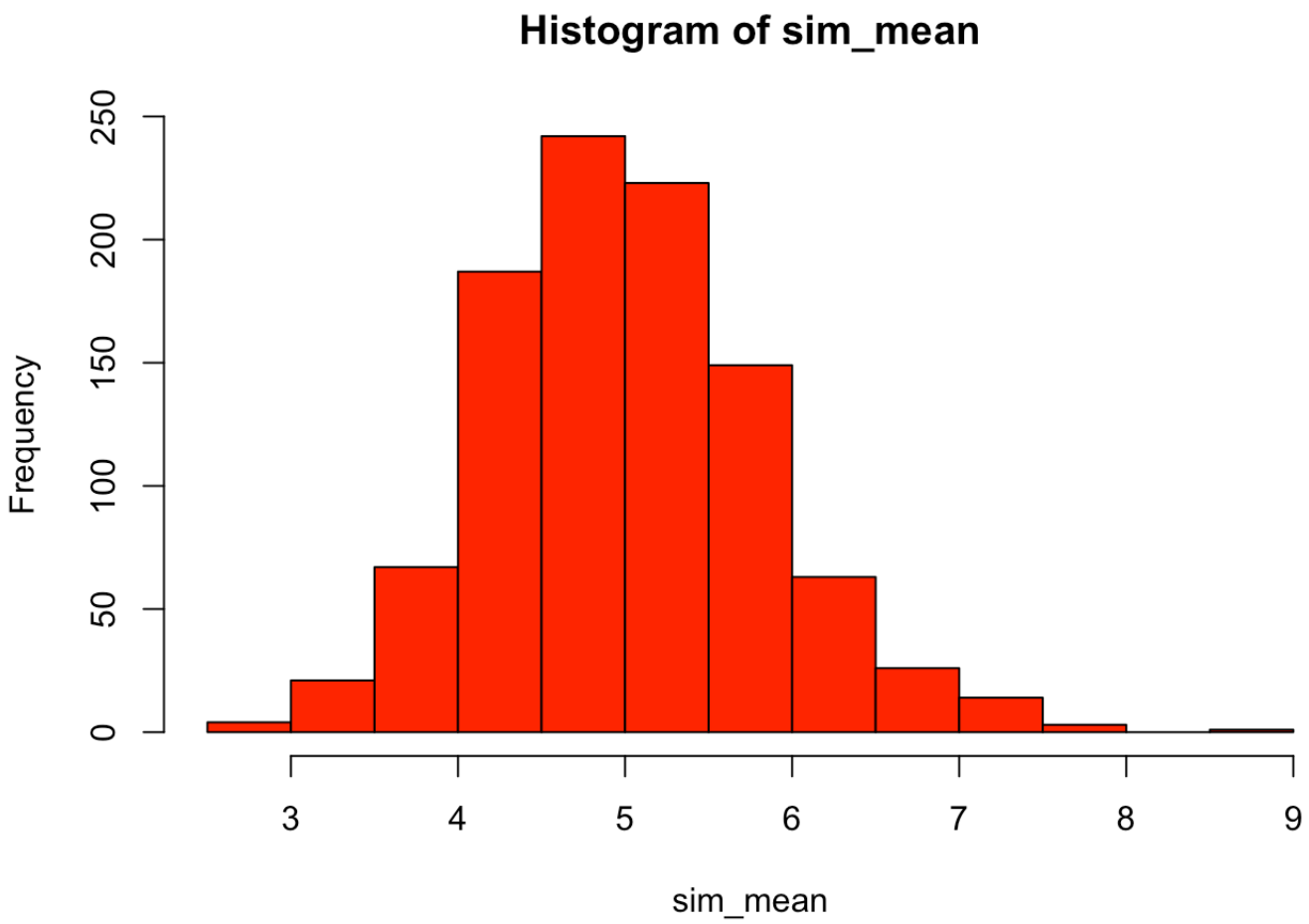
1. Show the sample mean and compare it to the theoretical mean of the distribution.
2. Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution.
3. Show that the distribution is approximately normal.

```
          lambda = 0.2
          n = 40
          nosim = 1000

          set.seed(234)

 #Create a matrix of 1000 rows with the columns corresponding to random simulation 40
times
sim_matrix <- matrix(rexp(nosim * n, rate=lambda), nosim, n)
sim_mean <- rowMeans(sim_matrix)

hist(sim_mean, col = "red")
```



Mean comparison

```
mean_data <- mean(sim_mean)
theory_mean <- 1/lambda

paste("sample mean is ", mean_data)
```

```
## [1] "sample mean is  5.0015728501858"
```

```
paste("theory mean is ", theory_mean)
```

```
## [1] "theory mean is  5"
```

Our sample mean is very close to theoretical mean

Variance comparison

```
actual_var <- var(sim_mean)
theory_var <- (1/lambda)^2/n

paste("sample variance is ", actual_var)
```

```
## [1] "sample variance is  0.66315043736661"
```

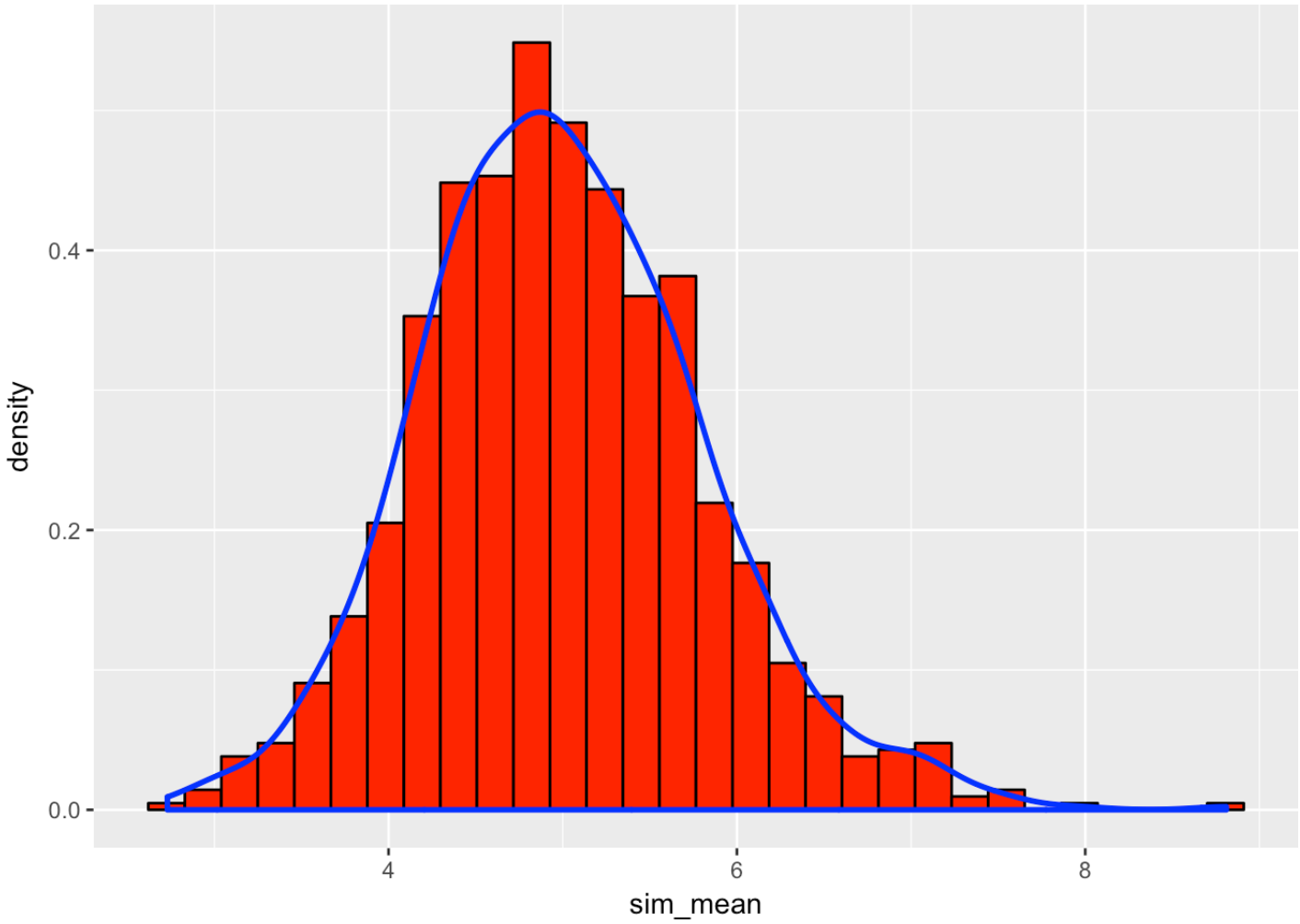
```
paste("theory variance is ", theory_var)
```

```
## [1] "theory variance is  0.625"
```

Normal Distribution

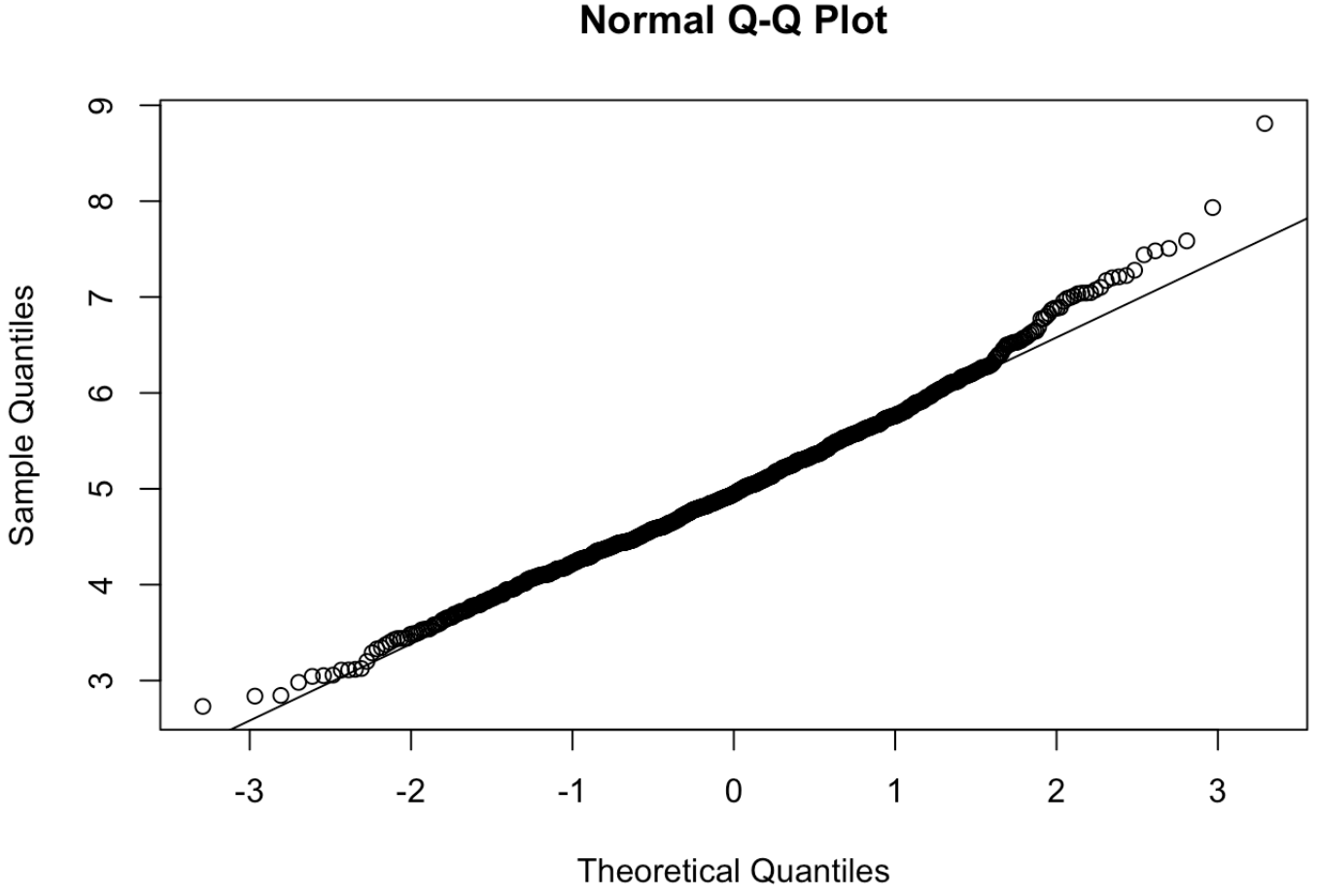
```
plotdata <- data.frame(sim_mean);
m <- ggplot(plotdata, aes(x =sim_mean))
m <- m + geom_histogram(aes(y=..density..), colour="black",
fill = "red")
m + geom_density(colour="blue", size=1);
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```




```
actual_conf_interval <- round (mean(sim_mean) + c(-1,1)*1.96*sd(sim_mean)/sqrt(n),3)
theory_conf_interval <- theory_mean + c(-1,1)*1.96*sqrt(theory_var)/sqrt(n);

qqnorm(sim_mean);
qqline(sim_mean)
```


Normal Q-Q Plot
Sample Quantiles
Theoretical Quantiles
3
4
5
6
7
8
9
-3
-2
-1
0
1
2
3